Pergamon

Computers Math. Applic. Vol. 35, No. 5, pp. 17–25, 1998

0898-1221/98 $19.00 + 0.00

PII: S0898-1221(98)00002-9

Optimal Control and Invexity

B. D. CRAVEN
Department of Mathematicsand Statistics, University of Melbourne
Parkville, Victoria 3052, Australia
craven@mundoe.maths.mu.oz.au

(Received May 1997; accepted June 1997)

Abstract—For a constrained minimization problem in infinite dimensions, in particular an optimal control problem, the attainment of a minimum follows if necessary Lagrangian conditions—Karush-Kuhn-Tucker or equivalently Pontryagin—are solvable, provided that a suitable invex hypothesis holds. Duality results are also obtained, where part of the constraint system describes a curved (hyper-) surface, and the invex property is assumed on that surface.

Keywords—Constrained minimization, Invex, Optimal control, Pontryagin, Duality.

1. INTRODUCTION

Consider a fixed-time optimal control problem:

$$\begin{aligned}
&\min_{\mathbf{x}(.),\mathbf{u}(.)} && J(\mathbf{u}) := F(\mathbf{x},\mathbf{u}) := \int_0^T f(\mathbf{x}(t),\mathbf{u}(t),t)\,dt, \\
&\text{subject to} && \mathbf{x}(0)=\mathbf{x}_0, \quad \left(\frac{d}{dt}\right)\mathbf{x}(t) = m(\mathbf{x}(t),\mathbf{u}(t),t), \qquad (0\le t\le T), \\
& && \mathbf{u}(t)\in\Gamma(t), \qquad (0\le t\le T).
\end{aligned} \tag{OC}$$

Here the control $\mathbf{u}(.)$ is assumed piecewise continuous, the state $\mathbf{x}(.)$ is assumed piecewise smooth, $f(.,.,.)$ and $m(.,.)$ are C^1 functions, and $\Gamma(t)$ is a given set, for each t. As in [1,2], the problem (OC) can be written as a mathematical program in function spaces, namely,

$$\min_{\mathbf{x},\mathbf{u}} F(\mathbf{x},\mathbf{u}), \qquad \text{subject to } D\mathbf{x} = M(\mathbf{x},\mathbf{u}), \qquad \mathbf{u}\in\Delta, \tag{OC2}$$

where $(\forall\, t)\ M(\mathbf{x},\mathbf{u})(t) := m(\mathbf{x}(t),\mathbf{u}(t),t)$, $\Delta := (\mathbf{u} : (\forall\, t)\ \mathbf{u}(t)\in\Gamma(t))$, and

$$\mathbf{y} = D\mathbf{x} \iff (\forall\, t)\ \mathbf{x}(t) = \mathbf{x}_0 + \int_0^t \mathbf{y}(s)\,ds.$$

With the norms $\|\mathbf{u}\|_\infty$ and $\|\mathbf{x}\| := \|\mathbf{x}\|_\infty + \|D\mathbf{x}\|_\infty$, $F(.,.)$ and $M(.,.)$ are Fréchet differentiable.

A differentiable vector function $\Xi(.)$, with gradient $\Xi'(.)$, is *invex* at a point p, with respect to a convex cone Q in its range space, if

$$(\forall\, z)\ \Xi(z) - \Xi(p) - \Xi'(p)\omega(z-p) \in Q, \tag{IX}$$

for some *scale function* $\omega(.)$ satisfying $\omega(z-p) = z - p + \mathrm{o}(\|z-p\|)$. This definition applies equally to functions on spaces of finite, or infinite, dimensions (see [2]). Consider the mathematical programming problem:

$$\min_z F(z), \qquad \text{subject to } -G(z)\in S, \qquad z\in\Delta, \tag{MP}$$

Typeset by $\mathcal{A}_\mathcal{M}\mathcal{S}$-TEX

with $F(.)$ and $G(.)$ differentiable, S a closed convex cone, Δ a closed set. Assume that a local minimum of (MP) is reached at $z = p$. It is well known that if $\Xi(.) := (F(.), G(.))$ is invex at p with respect to the convex cone $\mathbf{R}_+ \times S$, for some scale function $\omega(.)$ satisfying

$$\omega(z - p) \in \Delta - p,$$

then the necessary Karush-Kuhn-Tucker conditions (KKT) for a local minimum of (MP) become also sufficient for a local minimum of (MP)—even for a global minimum if the invex condition (IX) holds globally.

By considering the control problem (OC2) as a mathematical program (MP), it will be shown that invexity can provide also a condition under which a local minimum is reached. Also, when the variables of the problem are divided into two sets—state $x(.)$ and control $u(.)$ for a control problem, but also in some other cases—both the invex property, and a resulting dual problem, can be expressed in a different form, involving a costate variable. An equality constraint, such as $Dx = M(x, u)$, requires some special treatment, but it need not be linear (or affine), in order to construct a dual problem.

2. WHEN IS A MINIMUM REACHED?

Consider the constrained minimization problem (MP), with Z and W Banach spaces, $F : Z \to \mathbf{R}$ and $G : Z \to W$ (Fréchet) differentiable functions, and $S \subset W$ a closed convex cone. For control problems, Z and W are generally infinite-dimensional function spaces, so that compactness properties are generally lacking (in particular, the unit sphere is *not* compact), so that it is not a trivial question whether a minimum point is reached.

Suppose, however, that the problem (MP) reaches a local minimum at a point $p \in \Delta$, satisfying $-G(p) \in S$. Suppose also that the constraint satisfies Robinson's stability condition at p, namely that

$$0 \in \text{int}[G(p) + G'(p)(\Delta) + S],$$

where int denotes interior. Then [1,2], there hold necessary Karush-Kuhn-Tucker conditions:

$$F'(p) + \lambda G'(p) \in (\Gamma - p)^*, \qquad \lambda \in S^*, \qquad \lambda G(p) = 0. \tag{KKT}$$

Here $*$ denotes the positive dual cone, thus S^* is the set of dual vectors θ for which $\theta(S) \subset \mathbf{R}_+$.

Now, without assuming that a minimum is reached, suppose that (KKT) holds at a feasible point p. Then invexity will show that p is, in fact, a minimum point, and hence a minimum is reached.

THEOREM 1. *(See [2,3].) For the constrained minimization problem (MP), let $p \in \Delta$ satisfy $-G(p) \in S$, let $\Phi(.) := (F(.), G(.))$ be invex at p, with respect to the convex cone $\mathbf{R}_+ \times S$, and let $\omega(\Delta - p) \subset \Delta - p$. Then (MP) reaches a minimum at p.*

PROOF. Let $z \in \Delta$ and $-G(z) \in S$. From the invex hypothesis, with $\lambda \in S^*$, $(F + \lambda G)(.)$ is invex with respect to $\mathbf{R}_+$. Using KKT,

$$F(z) - F(p) \geq (F + \lambda G)(z) - (F + \lambda G)(p) \geq (F + \lambda G)'(p)\omega(z - p) \geq 0,$$

since (KKT) requires $(F + \lambda G)'(p) \in (\Delta - p)^*$, and $\omega(z - p) \in \Delta - p$. This (MP) reaches a minimum at p. ∎

This minimum point is local (or global), according to whether (IX) holds locally, for x near p (or globally). The invex hypothesis assumes more than is required for sufficiency, since (IX) applies to all z (or all z near p), whereas only those z satisfying the constraints are ever used. Moreover, invex is quite restrictive for an equality constraint, since $\omega(.)$ must satisfy an equation.

The constraint $-G(z) \in S$ can be transformed into other equivalent forms. Suppose that $\mathbf{T} : W \to W$ is a continuous linear mapping, such that $\mathbf{T}(S) = S$. Then

$$-G(z) \in S \iff -\mathbf{T}G(z) \in S.$$

Hence, it suffices if the invex property holds when $\Phi(.)$ is changed to $(F(.), \mathbf{T}G(.))$. In particular, if $W = \mathbf{R}^m$ and $S = \mathbf{R}^m_+$ (the nonnegative orthant), then any diagonal $m \times m$ matrix $\mathbf{T}$ with positive diagonal elements ν_j has the property $\mathbf{T}(S) = S$. The corresponding invex property is the *V-invex* property of Jeyakumar and Mond [4]. However, this property clearly extends to some polyhedral cones S, also to optimization in Hilbert spaces (supposing that $\{\nu_j\} \to 0$ suitably rapidly).

3. WHEN IS A CONTROL PROBLEM INVEX?

Consider the control problem (OC), assuming that F and M in (OC2) are twice-differentiable, and that $D\mathbf{x} = M(\mathbf{x}, \mathbf{u})$ defines $\mathbf{x}(.) = \Phi[\mathbf{u}](.)$, where Φ is Lipschitz. Define $q(\mathbf{u}(t), t) := f(\Phi[\mathbf{u}](t), \mathbf{u}(t), t)$. If $J(\mathbf{u}) := \int_0^T q(\mathbf{u}(t), t)\, dt$ is minimized over $\mathbf{u} \in \Delta$ at $\hat{\mathbf{u}}$, then the Karush-Kuhn-Tucker conditions (KKT) require that $J'(\hat{\mathbf{u}}) \in (\Gamma - \hat{\mathbf{u}})^*$, the dual cone of $\Gamma - \hat{\mathbf{u}}$. In particular, if the control constraint is $-G(\mathbf{u}) := -\,\mathbf{u}(.) \le 0$, and if (by suitable change of control variable) $\hat{\mathbf{u}} = 0$, then $J'(\hat{\mathbf{u}}) \ge 0$. Suppose now that $J(.)$ is invex; thus

$$J(\mathbf{u}) - J(\hat{\mathbf{u}}) \ge J'(\hat{\mathbf{u}})(\mathbf{u} - \hat{\mathbf{u}} + \theta(\mathbf{u} - \hat{\mathbf{u}})), \qquad \theta(\mathbf{u} - \hat{\mathbf{u}}) = \mathbf{o}(\|\mathbf{u} - \hat{\mathbf{u}}\|),$$

then Theorem 1 also requires that $-\mathbf{u} + \hat{\mathbf{u}} \ge (-1)(\mathbf{u} - \hat{\mathbf{u}} + \theta(\mathbf{u} - \hat{\mathbf{u}}))$, thus $\theta(\mathbf{u} - \hat{\mathbf{u}}) \ge 0$ in this example. With $J'(\hat{\mathbf{u}}) \ge 0$, the invexity of J reduces here to convexity.

Consider now a weaker control constraint as $\mathbf{u} \in \Delta$, with

$$\Delta = \{\mathbf{u} : (\forall t)\ 0 \le \mathbf{u}(t) \le \delta\}.$$

For a real variable z, $z + \mathbf{o}(z) \ge 0$ for sufficiently small $z \ge 0$. Hence, if

$$\omega(\mathbf{u})(t) = \mathbf{u}(t) + \frac{1}{2}\mathbf{u}(t)^T Q(t)\mathbf{u}(t)$$

(considering a quadratic θ), then the requirement $(\forall\, \mathbf{u} \ge 0)\ \omega(\mathbf{u})(t) \in [0, \delta]$ is fulfilled for all t if $Q(t)$ is a diagonal matrix for which $(\forall t)\ \|Q(t)\| \le \delta'$, for suitable δ' depending on δ. Note that the quadratic term is $\mathbf{o}(\mathbf{u}(t))$, not merely $\mathbf{o}(\mathbf{u}(.))$, since the times t and components of $\mathbf{u}$ must be considered separately. Denote $\mathbf{v} := \mathbf{u} - \hat{\mathbf{u}}$. Now, make the *pointwise invex* hypothesis that

$$\begin{gathered}(\forall t)\ q(\mathbf{u}(t), t) - q(\mathbf{u}^*(t), t) \ge \sigma(t)\mathbf{v}(t) + \frac{1}{2}\mathbf{v}(t)^T M(t)\mathbf{v}(t), \\ \|Q(t)\| \le \delta'. \end{gathered} \tag{PInv}$$

Then $J(.)$ is invex, and $(\forall\, \mathbf{u} \ge 0)\ \omega(\mathbf{u})(t) \in [0, \delta]$. From Theorem 1, KKT is then necessary for a local minimum of (OC).

THEOREM 2. *For the control problem (OC) with control constraint*

$$\mathbf{u}(t) \in [a, b] \subset \mathbf{R},$$

assume (PInv) *at* $\hat{\mathbf{u}}$, *where* $\hat{\mathbf{u}}(.) \in [a, b]$; *let* $\delta > 0$ *and* $\delta' > 0$ *satisfy* $\delta'\delta < 2$, *and* $(\forall t)\ \delta < b - a$ *if* $\hat{\mathbf{u}}(t) = a$ *or* b, $\delta < \min\{\hat{\mathbf{u}}(t) - a, b - \hat{\mathbf{u}}(t)\}$ *otherwise. Then* $J(.)$ *is invex, and* $\omega(\mathbf{u}(t) - \hat{\mathbf{u}}(t)) \in [a, b] - \hat{\mathbf{u}}(t)$ *whenever* $\hat{\mathbf{u}}(t) \in [a, b]$ *and* $|\mathbf{u}(t) - \hat{\mathbf{u}}(t)| < \delta$. *So, the hypotheses of Theorem 1 are satisfied.*

PROOF. Denote $z = \mathbf{u}(t) - \hat{\mathbf{u}}(t)$. For $|z| \le \delta$,

$$\hat{\mathbf{u}}(t) + z \in [a, b] \Rightarrow z + \frac{1}{2}Qz^2 \in \operatorname{sgn}(z).[0, 2] \subset [a, b] - \hat{\mathbf{u}}(t). \qquad \blacksquare$$

4. PARTIAL DUALITY UNDER AN INVEX HYPOTHESIS

Consider now a constrained minimization problem, with two vector constraints, which will be treated differently:

$$\min_{x,y} F(x,y), \qquad \text{subject to } -K(x,y) \in T, \qquad -G(x,y) \in S. \tag{MP2}$$

Assume that $F(.,.)$ and $G(.,.)$ are Fréchet differentiable, $x \in X_1$ and $y \in X_2$, where X_1 and X_2 are normed spaces, and T and S are closed convex cones, in normed spaces Z_1 and Z_2, respectively.

The constraints can be rewritten, with a slack variable t, as

$$K(x,y) + t = 0, \qquad t \in T, \qquad -G(x,y) \in S.$$

Thus the constraints of (MP2) become

$$\tilde{K}(x,y,t) := K(x,y) + t = 0, \qquad -\tilde{G}(x,y,t) := -[G(x,y), -t] \in S \times T,$$

thus with y extended to (y,t), and an equality constraint $\tilde{K}(x,y,t) = 0$.

So, it suffices to discuss an equality constraint $K(x,y) = 0$. Hence, consider now the problem:

$$\min_{x,y} F(x,y), \qquad \text{subject to } K(x,y) = 0, \qquad -G(x,y) \in S, \tag{MP3}$$

subject to the following hypotheses.

HYPOTHESIS S. *The equation $K(x,y)$ is solvable uniquely for $x = \Xi(y)$, where $\Xi(.)$ is Lipschitz.*

HYPOTHESIS U. *For all $x := \Xi(y)$ and $\hat{x} := \Xi(\hat{y})$,*

$$F(x,u) - F(\hat{x},u) = F_x(\hat{x},\hat{u})(x-\hat{x}) + \mathbf{o}(\|x-\hat{x}\| + \|u-\hat{u}\|),$$
$$K(x,u) - K(\hat{x},u) = K_x(\hat{x},\hat{u})(x-\hat{x}) + \mathbf{o}(\|x-\hat{x}\| + \|u-\hat{u}\|),$$

where F_x and K_x denote partial derivatives with respect to x.

Hypothesis S may be considered as holding locally, near a point $(\hat{x},\hat{y}) = p$ where $K(\hat{x},\hat{y}) = 0$, or holding globally, for unrestricted y. The first case corresponds to applying an implicit function theorem (such as Lyusternik's) to solve for $x = \Xi(y)$; the second case corresponds to the typical control problem, where $K(x,y) = 0$ represents a differential equation for x in terms of the control y, and a global solution may be available. Hypothesis U (see [2, Theorem 7.2.3]) may be considered as partial differentiability with respect to x, uniformly with respect to u near $\hat{u}$. Note that, if $\Xi(.)$ is Lipschitz then

$$\mathbf{o}(\|x-\hat{x}\| + \|u-\hat{u}\|) = \mathbf{o}(\|u-\hat{u}\|). \tag{Lip}$$

Define $J(y) := F(\Xi(y), y)$. For λ in the dual space of Z_1, define

$$H(x,y,\lambda) := F(x,y) + \lambda K(x,y).$$

THEOREM 3. *For problem (MP3), assume that $F(.,y)$ and $K(.,y)$ are Fréchet differentiable; assume Hypotheses S and U; assume that $\hat{\lambda}$ satisfies $H_x(\hat{x},\hat{y},\hat{\lambda}) = 0$. Then the gradient*

$$J_y(\hat{y}) = H_y\left(\hat{x},\hat{y},\hat{\lambda}\right).$$

PROOF. Let $x := \Xi(y)$ and let $\hat{x} := \Xi(\hat{y})$. Then

$$H\left(\hat{x},\hat{y},\hat{\lambda}\right) - H\left(\hat{x},y,\hat{\lambda}\right) = F(\hat{x},\hat{y}) - F(x,y) + F(x,y) - F(\hat{x},y) + \hat{\lambda}[K(\hat{x},\hat{y}) - K(\hat{x},y)],$$

$$
\begin{aligned}
&= F(\hat{x},\hat{y}) - F(x,y) + F(x,y) - F(\hat{x},y) + \hat{\lambda}[K(x,y) - K(\hat{x},y)] \\
&\quad \text{(since } K(x,y) = K(\hat{x},\hat{y}) = 0\text{)}, \\
&= F(\hat{x},\hat{y}) - F(x,y) + H\left(x,y,\hat{\lambda}\right) - H\left(\hat{x},y,\hat{\lambda}\right) \\
&= F(\hat{x},\hat{y}) - F(x,y) + H_x\left(\hat{x},\hat{y},\hat{\lambda}\right)(x-\hat{x}) + \mathbf{o}(\|x-\hat{x}\| + \|y-\hat{y}\|) \\
&= F(\hat{x},\hat{y}) - F(x,y) + 0 + \mathbf{o}(\|y-\hat{y}\|) \qquad \text{from (Lip)} \\
&= J(\hat{y}) - J(y) + \mathbf{o}(\|y-\hat{y}\|),
\end{aligned}
$$

proving the result. ∎

The problem (MP3) can be expressed as a two-stage minimization:

$$\min_y\{\min_x F(x,y),\ \text{subject to } K(x,y)=0\}, \qquad \text{subject to } -G(x,y)\in S.$$

The inner minimization is here trivial, given Hypothesis S, so that the outer problem becomes:

$$\min_y J(y) := F(\Xi(y),y), \qquad \text{subject to } -P(y) := -G(\Xi(y),y) \in S. \tag{MP4}$$

Let $\Psi(.) := (J(.), P(.))$. If the hypotheses of Theorem 2 hold, then $\Psi(.)$ is invex at $(\hat{x},\hat{y})$, with respect to the convex cone $\mathbf{R}_+ \times S$, with scale function $\omega(.)$, exactly when

$$F(x,y) - F(\hat{x},\hat{y}) - H_y\left(\hat{x},\hat{y},\hat{\lambda}\right)\omega(y-\hat{y}) \in \mathbf{R}_+ \times S,$$

subject to $K(x,y) = 0 = K(\hat{x},\hat{y})$, and $\hat{\lambda}$ satisfying $H_x(\hat{x},\hat{y},\hat{\lambda}) = 0$.

Assuming a constraint qualification, necessary KKT conditions for a local minimum of (MP3) at $(\hat{x},\hat{y})$ are:

$$J_y(\hat{y}) + \rho P_y(\hat{y}) = 0, \qquad \rho \in S^*, \qquad \rho P(\hat{y}) = 0. \tag{KKT1}$$

Applying Theorem 3 to F, K, and J with multiplier λ, and also to G, K, and P with a new multiplier σ, (KKT1) is equivalent to:

$$(F+\lambda K)_y(\hat{x},\hat{y}) + \rho(G+\sigma K)_y(\hat{x},\hat{y}) = 0, \qquad \rho\in S^*, \qquad \rho G(\hat{x},\hat{y}) = 0, \tag{KKT2}$$

subject, however, to the side conditions

$$K(\hat{x},\hat{y}) = 0, \qquad (F+\lambda K)_x(\hat{x},\hat{y}) = 0, \qquad (G+\sigma K)_x(\hat{x},\hat{y}) = 0. \tag{C}$$

This has proved the following result.

THEOREM 4. *For the problem (MP3), assume Hypothesis S, Hypothesis U, and a similar hypothesis to U for $G(.,.)$ replacing $F(.,.)$. Assume that multipliers λ and σ exist, satisfying (C). Assume the Robinson stability condition for G. Then necessary conditions for a local minimum of (MP3) at $(\hat{x},\hat{y})$ are that (KKT2) holds, subject to (C).*

REMARK. A *Mond-Weir dual* problem to (MP2) can now be formulated:

$$\max_{v,\rho}(J(v) + \rho P(v)), \qquad \text{subject to } J_y(v) + \rho P_y(v) = 0, \quad \rho\in S^*, \quad \rho P(v) \geq 0.$$

This is equivalent, under the hypotheses of Theorem 4, to

$$
\begin{aligned}
\max_{\xi,\eta,\rho,\lambda,\sigma}\quad & (F(\xi,\eta) + \rho G(\xi,\eta)), \\
\text{subject to}\quad & (F+\lambda K)_y(\xi,\eta) + \rho(G+\sigma K)_y(\xi,\eta) = 0, \quad \rho\in S^*, \quad \rho G(\xi,\eta) = 0, \\
& K(\xi,\eta) = 0, \quad (F+\lambda K)_x(\xi,\eta) = 0, \quad (G+\sigma K)_x(\xi,\eta) = 0, \quad \sigma = S^*.
\end{aligned}
\tag{MD2}
$$

Then (MD2) is, in fact, a dual problem under invex hypotheses, as follows.

HYPOTHESIS IX2. $(\exists\,\omega : X_2 \times X_2 \to X_2, \omega(0,.) = 0, \omega(\zeta,.) = \zeta + \mathbf{o}(\zeta))$, $(\exists\,\lambda, (F+\lambda L)_x(\xi,\eta) = 0)$, $(\exists\,\sigma, (G+\sigma K)_x(\xi,\eta) = 0)$, $(\forall\, x, y : K(x,y) = 0)$, $(\forall\, \xi, \eta, K(\xi,\eta) = 0)$,

$$F(x,y) - F(\xi,\eta) - (F+\lambda K)_y(\xi,\eta)\omega(y-\eta,\eta) \geq 0,$$
$$G(x,y) - G(\xi,\eta) - (G+\sigma K)_y(\xi,\eta)\omega(y-\eta,\eta) \in S.$$

THEOREM 5. *Assume that problem (MP2) satisfies the hypotheses of Theorem 3, and also the invex Hypothesis IX2. Then (MD2) is a strong dual problem to (MP2).*

PROOF. In view of Theorem 4, it is enough to show that (MD2) is a strong dual to (MP4). There is zero duality gap, from Theorem 3. From Hypotheses I, substituting the gradient of $J(.)$ from Theorem 2 into the Definition IX of *invex* for $J(.)$, there follows

$$J(y) - J(\eta) \geq J_y(\eta)\omega(y-\eta,\eta),$$

and similarly

$$P(y) - P(\eta) - P_y(\eta)\omega(y-\eta,\eta) \in S,$$

whenever $K(x,y) = 0$ and $K(\xi,\eta) = 0$. Thus $(J(.), P(.))$ is invex, at each point η, with respect to the convex cone $\mathbf{R}_+ \times S$, on the restricted domain of points satisfying $K(.,.) = 0$. Suppose now that y is feasible for (MP4), and (v,w) is feasible for (MD2). Then

$$J(y) - J(\eta) \geq (F+wP)(y) - (F+wP)(\eta) \geq (F+wP)_y(v)\omega(y-v,v) = 0,$$

proving weak duality. ∎

REMARK. This is the standard proof of Mond-Weir weak duality, but with the invex property holding on the stated restricted domain.

REMARK. Consider the control problem (OC), with the control constraint $(\forall\, t)\ \mathbf{u}(t) \in \Gamma(t)$ replaced by a constraint

$$(\forall\, t)\ g(\mathbf{x}(t), \mathbf{u}(t), t) \leq 0.$$

The Mond-Weir dual takes the form:

$$\begin{aligned}
\max \quad & \int_0^T [f(\xi(t),\zeta(t),t) + \nu(t)g(\xi(t),\zeta(t),t)]\,dt,\\
\text{subject to} \quad & \xi(0) = \mathbf{x}_0, \left(\frac{d}{dt}\right)\xi(t) = m(\xi(t),\zeta(t),t); \quad \nu(t) \geq 0, \quad \nu(t)g(\xi(t),\zeta(t),t) \geq 0,\\
& -D\lambda(t) = (f+\lambda(t))m(\xi(t),\zeta(t)), \quad \lambda(T) = 0,\\
& -D\sigma(t) = (g+\sigma(t))(\xi(t),\zeta(t),t), \quad \sigma(T) = 0, \quad \sigma(t) \in S^*.
\end{aligned}$$

Here $\nu(t), \lambda(t), \sigma(t)$ are functions representing the Lagrange multipliers ρ, λ, and σ. The dual constraints for $-D\lambda(t)$ and $-D\sigma(t)$ are two adjoint differential equations.

5. SECOND-ORDER SENSITIVITY

Consider the optimal control problem (OC2), with state $\mathbf{x}$ and control $\mathbf{u}$ (and no state constraint). Denote

$$P(\mathbf{x},\mathbf{u}) \equiv -D\mathbf{x} + M(\mathbf{x},\mathbf{u}), \qquad Q(\mathbf{x},\mathbf{u};\boldsymbol{\lambda}) := F(\mathbf{x},\mathbf{u}) + \boldsymbol{\lambda}P(\mathbf{x},\mathbf{u}),$$
$$H(\mathbf{x},\mathbf{u};\boldsymbol{\lambda}) := F(\mathbf{x},\mathbf{u}) + \boldsymbol{\lambda}M(\mathbf{x},\mathbf{u}).$$

Then

$$H(\mathbf{x},\mathbf{u};\boldsymbol{\lambda}) = \int_0^T h(\mathbf{x}(t),\mathbf{u}(t),t;\lambda(t))\,dt,$$

where $\lambda(.)$ is the costate, and the Hamiltonian is

$$h(\mathbf{x}(t), \mathbf{u}(t), t; \lambda(t)) = f(\mathbf{x}(t), \mathbf{u}(t), t) + \lambda(t) m(\mathbf{x}(t), \mathbf{u}(t), t).$$

Assume F and M are twice-differentiable, and that $D\mathbf{x} = M(\mathbf{x}, \mathbf{u})$ defines $\mathbf{x}(.) = \Phi[\mathbf{u}(.)]$, where Φ is Lipschitz. For brevity, denote $\mathbf{x} = \Phi[\mathbf{u}]$ and $\hat{\mathbf{x}} = \Phi(\hat{\mathbf{u}})$, for two controls $\mathbf{u}$ and $\hat{\mathbf{u}}$. Under some regularity (see [2, Theorem 7.2.3]), calculating up to second-order terms, with $\hat{\boldsymbol{\lambda}}$ satisfying $Q_{\mathbf{x}}(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}) = 0$,

$$\begin{aligned} H\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right) - H\left(\hat{\mathbf{x}}, \mathbf{u}; \hat{\boldsymbol{\lambda}}\right) &= F(\hat{\mathbf{x}}, \hat{\mathbf{u}}) - F(\mathbf{x}, \mathbf{u}) + Q\left(\mathbf{x}, \mathbf{u}; \hat{\boldsymbol{\lambda}}\right) - Q\left(\hat{\mathbf{x}}, \mathbf{u}; \hat{\boldsymbol{\lambda}}\right) \\ &= F(\hat{\mathbf{x}}, \hat{\mathbf{u}}) - F(\mathbf{x}, \mathbf{u}) + Q_{\mathbf{x}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right)(\mathbf{x} - \hat{\mathbf{x}}) \\ &\quad + \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^T Q_{\mathbf{xx}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right)(\mathbf{x} - \hat{\mathbf{x}}) \\ &\quad + \left[Q_{\mathbf{x}}\left(\hat{\mathbf{x}}, \mathbf{u}; \hat{\boldsymbol{\lambda}}\right) - Q_{\mathbf{x}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right)\right](\mathbf{u} - \hat{\mathbf{u}}) \qquad \text{(Ham)} \\ &= F(\hat{\mathbf{x}}, \hat{\mathbf{u}}) - F(\mathbf{x}, \mathbf{u}) + Q_{\mathbf{x}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right)(\mathbf{x} - \hat{\mathbf{x}}) \\ &\quad + \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^T Q_{\mathbf{xx}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right)(\mathbf{x} - \hat{\mathbf{x}}) \\ &\quad + (\mathbf{x} - \hat{\mathbf{x}})^T Q_{\mathbf{xu}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right)(\mathbf{u} - \hat{\mathbf{u}}), \end{aligned}$$

up to second-order terms, where

$$(\mathbf{x} - \hat{\mathbf{x}})^T Q_{\mathbf{xu}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right)(\mathbf{u} - \hat{\mathbf{u}}) = \mathrm{o}(\|\mathbf{u} - \hat{\mathbf{u}}\|).$$

Assume that the linearized equation

$$D(\mathbf{x} - \hat{\mathbf{x}}) = M_{\mathbf{x}}(\hat{\mathbf{x}}, \hat{\mathbf{u}})(\mathbf{x} - \hat{\mathbf{x}}) + M_{\mathbf{u}}(\hat{\mathbf{x}}, \hat{\mathbf{u}})(\mathbf{u} - \hat{\mathbf{u}})$$

defines $\mathbf{x} - \hat{\mathbf{x}} = K(\mathbf{u} - \hat{\mathbf{u}})$, where K is a linear mapping. Neglecting terms in (Ham) higher than second-order, $\mathbf{x} - \hat{\mathbf{x}}$ may be replaced there by $K(\mathbf{u} - \hat{\mathbf{u}})$. Hence, to this approximation, and writing $\mathbf{v} := \mathbf{u} - \hat{\mathbf{u}}$,

$$\begin{aligned} F(\mathbf{x}, \mathbf{u}) - F(\hat{\mathbf{x}}, \hat{\mathbf{u}}) &= H_{\mathbf{u}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right)\mathbf{v} + \frac{1}{2}\mathbf{v}^T H_{\mathbf{uu}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right)\mathbf{v} \\ &\quad + \frac{1}{2}\mathbf{v}^T K^T Q_{\mathbf{xx}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right)K\mathbf{v} + \mathbf{v}^T K^T Q_{\mathbf{xu}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right)\mathbf{v} \qquad \text{(Quad)} \\ &= H_{\mathbf{u}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right)\mathbf{v} + \frac{1}{2}\mathbf{v}^T R^{\#}\mathbf{v}, \end{aligned}$$

where

$$R^{\#} = H_{\mathbf{uu}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right) + K^T H_{\mathbf{xx}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right)K + 2K^T H_{\mathbf{xu}}\left(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\lambda}}\right).$$

Equivalently, $R^{\#}$ may be replaced by the symmetric matrix

$$R := \begin{bmatrix} K^T & I \end{bmatrix} \begin{bmatrix} H_{\mathbf{xx}} & H_{\mathbf{xu}} \\ H_{\mathbf{ux}} & H_{\mathbf{uu}} \end{bmatrix} \begin{bmatrix} K \\ I \end{bmatrix}.$$

Note, however, that K is not pointwise operator. From $D\mathbf{x} = M(\mathbf{x}, \mathbf{u})$,

$$\mathbf{x}(t) = (K\mathbf{u})(t) = \int_0^T k(t, s)\mathbf{u}(s)\, ds,$$

for some kernel $k(.,.)$ satisfying $k(s,t)=0$ for $s>t$. For brevity, write

$$\left(\begin{bmatrix} K \\ I \end{bmatrix} \mathbf{u}\right)(t) := \int_0^T \hat{k}(t,s)\mathbf{u}(s)\,ds,$$

where the kernel $\hat{k}$ comprises k and a delta-function term. Denote the Hamiltonian for (OC1) by

$$h(\mathbf{x}(t),\mathbf{u}(t),t;\lambda(t)) := f(\mathbf{x}(t),\mathbf{u}(t),t) + \lambda(t)m(\mathbf{x}(t),\mathbf{u}(t),t),$$

where $\lambda(.)$ is a costate function, and define the matrix function

$$r(t) := \begin{bmatrix} h_{\mathbf{xx}}(\mathbf{x}(t),\mathbf{u}(t),t;\lambda(t)) & h_{\mathbf{xu}}(\mathbf{x}(t),\mathbf{u}(t),t;\lambda(t)) \\ h_{\mathbf{ux}}(\mathbf{x}(t),\mathbf{u}(t),t;\lambda(t)) & h_{\mathbf{uu}}(\mathbf{x}(t),\mathbf{u}(t),t;\lambda(t)) \end{bmatrix}.$$

Then

$$\mathbf{v}^T R^{\#}\mathbf{v} = \mathbf{v}^T R\mathbf{v} = \int_0^T \int_0^T \mathbf{v}(s)^T W(s,s')\mathbf{v}(s')\,ds\,ds',$$

where

$$W(s,s') := \int_0^T \hat{k}(t,s')^T r(t)\hat{k}(t,s)\,dt.$$

In particular, consider $\mathbf{v}(t)=\mathbf{z}$, a constant, for $t\in(t_0-\delta,t_0+\delta)\subset[0,T]$, and zero elsewhere. If appropriate continuity is assumed, then $\delta^{-2}\mathbf{v}^T R\mathbf{v}$ reduces to $\mathbf{z}^T w(t_0)\mathbf{z}+o(\delta)$. With $q(\mathbf{u}(t),t) := f(\Phi[\mathbf{u}](t),\mathbf{u}(t),t)$, $\sigma(t) := \left(\frac{\partial}{\partial \mathbf{x}}\right) h(\mathbf{x}(t),\mathbf{u}(t),t;\lambda(t))$, and $w(t) := W(t,t)$,

$$\begin{aligned} q(\hat{\mathbf{u}}(t_0)+\mathbf{v}(t_0),t_0) - q(\hat{\mathbf{u}}(t_0),t_0) &= J(\hat{\mathbf{u}}+\mathbf{v}) - J(\hat{\mathbf{u}}) \\ &= \sigma(t_0)\mathbf{z} + \mathbf{z}^T w(t_0)\mathbf{z} + o(\delta^2), \end{aligned}$$

assuming continuity. Hence q has gradient and Hessian given by

$$q_{\mathbf{u}}(\hat{\mathbf{u}}(t),t) = \sigma(t) \qquad \text{and} \qquad q_{\mathbf{uu}}(\hat{\mathbf{u}}(t),t) = w(t).$$

Conditions necessary for invexity of $q(.,t)$, as sought in Section 3, are obtained by expansion of $q(.,t)$ and $\omega(.)$ up to quadratic terms. The necessary condition is then

$$w(t) - \sigma(t)M(t) \geq 0, \tag{Inv?}$$

where ≥ 0 for matrices means positive semidefinite. This condition will be locally sufficient for invexity, if the higher order terms happen to be dominated by the quadratic terms. In particular, if $\mathbf{u}(t)\in\mathbf{R}$, and if $w(t)$ has a negative eigenvalue $\nu(t)$, then the boundedness condition placed on $M(t)$ in Section 3 requires that $\nu(t)\to 0$ as $t\to t^\circ$, for any point t° where $\sigma(t^\circ)=0$.

6. FINITE-DIMENSIONAL APPROXIMATION

In order for the Pontryagin theory to apply to (OC), the minimum must still be a (local or global) minimum when the space of controls $\mathbf{u}$ has the $L^1(0,T)$ norm. Assuming that $J(\mathbf{u}) := F(\Phi[\mathbf{u}],\mathbf{u})$ is bounded below on Δ, an infimum exists, but a minimum is not necessarily reached.

Now consider the problem (OC-U), obtained from (OC) by restricting $\mathbf{u}(.)$ to a finite-dimensional subspace U. Then (OC-U) reaches a minimum, say, at $\mathbf{u}=\mathbf{u}^U\in U$; denote by $\mathbf{x}^U$ the corresponding state. Then KKT applies, with gradients with respect to $\mathbf{u}$ restricted to the subspace U. The part of KKT concerning gradient with respect to $\mathbf{x}$ leads to the adjoint differential equation, as for (OC). The calculation of the gradient $J'(\mathbf{u})$ in terms of the Hamiltonian is the same as for (OC), except that the gradient is now restricted to U.

Consider now a particular case, where $\Gamma(t) = [0, k]$, U is a space of step-functions which are constant on intervals I_j, and the Hamiltonian is linear in $\mathbf{u}$, say with a term $\sigma^U(t)\mathbf{u}(t)$. Then, the gradient satisfies

$$J'(\mathbf{u})|_U \mathbf{z} = \int_0^T \sigma^U(t)\mathbf{z}(t),$$

and $\sigma^U(t)\mathbf{u}(t)$ is minimized at $\mathbf{u}(t) = \mathbf{u}^U(t)$. The possible cases are:

$$\begin{array}{lll} \text{(i)} & \mathbf{u}^U(t) = 0, & \int_{I_j} \sigma^U(t)\,dt \geq 0, \\ \text{(ii)} & \mathbf{u}^U(t) = k, & \int_{I_j} \sigma^U(t)\,dt \leq 0, \\ \text{(iii)} & 0 < \mathbf{u}^U(t) < k, & \int_{I_j} \sigma^U(t)\,dt = 0. \end{array}$$

Consider now a sequence $\{U(j)\}$, where $U(j)$ is the subspace of step-functions constant on subintervals of length 2^{-j}, for $j = 1, 2, \ldots$. For case (i) above, $\{\mathbf{u}^{U(j)}\} \to \mathbf{u}(.) = 0$. Assuming some smoothness for the differential equations (bounded second derivatives will suffice), a subsequence of $\{\sigma^{U(j)}(.)\}$ converges, say, to $\sigma(.) \geq 0$, and the corresponding Hamiltonian term for problem (OC) is $\sigma(t)\mathbf{u}(t)$. Of course, Pontryagin theory predicts this, if a minimum is known to be attained, but that assumption was not required here.

REFERENCES

1. B.D. Craven, *Mathematical Programming and Control Theory*, Chapman and Hall, London, (1978).
2. B.D. Craven, *Control and Optimization*, Chapman and Hall, London, (1995).
3. M.A. Hanson, On the sufficiency of the Kuhn-Tucker conditions, *J. Math. Anal. Optim.* **8**, 545–550 (1980).
4. V. Jeyakumar and B. Mond, On generalized convex mathematical programming, *J. Austral. Math. Soc. Series B* **34**, 43–53 (1992).